在非高斯状态空间模型下对SCD模型的估计

靳　珊[1],黄荣坦[2]
(1.集美大学 理学院,福建 厦门 361006; 2.厦门大学 数学科学学院,福建 厦门 361021)

摘　要: 高频超高频时间序列的分析与建模成已为计量经济的一个全新研究领域,而研究金融市场中交易事件到达时间的随机条件持续期SCD模型,因为加入了随机变量,可以更好地拟合高频超高频金融时间序列特有的统计特征,但随机变量的引入给模型估计带来了估计困难。考虑到非高斯状态空间模型与随机条件持续期SCD模型各自的优势,文章将SCD模型转换成非高斯状态空间模型,从而利用非高斯状态空间框架下的Kalman滤波解决了SCD模型的估计难题。

关键词: 状态空间;非高斯;SCD模型
中图分类号: O212.7　　**文献标识码:** A　　**文章编号:** 1002-6487(2011)15-0017-04

0 引言

在传统经济学的观点里,短期波动是不相关的白噪声因而不值得收集,但是近期金融领域的研究改变了这样的观点,大家逐渐认同高频超高频数据中包含了大量有用信息并且能表现出长期趋势。我们知道金融市场中信息是连续地影响证券市场价格运动过程的,数据的离散采集必然会造成信息不同程度的缺失。无疑,信息采集频率越高,信息丢失越少;反之,信息丢失越多。所以,高频时间序列比低频时间序列包含更多的信息,如市场微观结构的信息及重要的长期日间现象的信息等。由于高频超高频时间序列数据时间间融的随机性,传统的时间序列建模方法不再适合于高频超高频时间序列。因此,人们试图建立新的计量经济模型来刻画此类数据。

引用Engle的话,对高频超高频时间序列的研究应该建立在对持续期建模的基础上进行。目前对持续期建模主要有两类,一类是Engle和Rusell(1998)提出的用随机标值点过程来刻画交易过程的ACD类模型[1],经过近几年的发展,已经形成了一个庞大的ACD模型族。另一类是Bauwenst和Veredas(2004)提出的带双随机过程的SCD类模型[2]。该模型假设存在一个产生持续期的潜在随机变量。他们认为该随机变量能捕捉到金融市场上的随机信息流,而这种信息流通常是不能直接观察到的,且引入的随机变量可以更好的拟合高频超高频金融时间序列特有的统计特征,已被国内外学者证实比ACD模型具有更好的拟合优度。SCD模型与ACD模型最主要的区别是SCD模型是双随机过程,即模型具有两个随机信息:一个针对持续期本身而言,即随机扰动项;另一个是针对潜在变量,即假设潜在变量是随机变量。也就是说,ACD模型中的条件期望持续期变成了SCD模型中的随机变量。这种改进类似于SV对GARCH类模型的改进。虽SCD模型已被提出,但同SV模型一样,由于似然函数中包含着一个不可观测的隐含变量,参数估计比较困难,一般的似然函数方法很难实现。对于SCD模型的估计,可采用一般矩估计GMM,有效矩估计EMM或基于Monte Carlo Markov chain(MCMC)的贝叶斯估计[3],模型的提出者Bauwens在最初是利用经典的状态空间模型进行QML估计,但经典的状态空间模型有着高斯假设的前提,估计并非有效。随着非高斯状态空间模型理论的进一步发展及重要抽样技术的不断改进,本文拟考虑利用非高斯状态空间模型对SCD模型进行估计并与Bauwens的QML估计进行比较,从模拟数据及实例数据两方面来说明估计的有效性得到提高,标准差都较小,采用的估计方法在计算速度上也有较大改善。本文与相关研究不同的地方是对独立样本进行抽样而没有使用马尔可夫链,这样做避免讨论收敛性问题。并且在非高斯状态模型中对状态方程和观测方程中的非高斯分布都将进行考虑,主要考虑的是经济时间序列中常出现的指数族分布如Poisson分布和重尾分布如t分布、混合高斯分布等。

1 模型简介

1.1 SCD模型[2]

SCD模型假设存在一个产生持续期的潜在随机变量,定义 x_i 为一个事件在时刻 t_{i-1} 发生到时刻 t_i 发生的持续期,为建立持续期过程的相依性,对数条件均值设定为服从一平稳的AR(1)过程。简单的SCD(1,1)模型为:

$$\begin{cases} x_i = e^{\psi_i}\varepsilon_i \\ \psi_i = \omega + \beta\psi_{i-1} + u_i(|\beta| < 1) \end{cases} \tag{1}$$

这里可设 $u_i|F_{i-1} \sim N(0, \sigma^2)$, $\varepsilon_i|F_{i-1}$ 服从某个带正支撑

作者简介: 靳　珊(1976-),女,山东人,硕士,研究方向:时间序列、计量经济。

的分布 $p(\varepsilon_i)$，且两个随机变量相互独立(其中 F_{t-1} 表示过去直到t-1时的信息集)。 $|\beta|<1$ 保证了平稳性，且 $|\beta|$ 的值越接近于1说明具有越高的集聚性和越强的持续性。其矩和自相关函数分别为：

$$\begin{cases} E(x_i)=\mu\exp\left(\dfrac{\omega}{1-\beta}+\dfrac{1}{2}\dfrac{\sigma^2}{1-\beta^2}\right) \\ \sigma_\psi^2=\exp\left(\dfrac{2\omega}{1-\beta}+\dfrac{\sigma^2}{1-\beta^2}\right)\left\{\exp\left(\dfrac{\sigma^2}{1-\beta^2}\right)\right\}-1 \\ \sigma_{xi}^2=\mu^2\exp\left(\dfrac{2\omega}{1-\beta}+\dfrac{\sigma^2}{1-\beta^2}\right)\left\{k\exp\left(\dfrac{\sigma^2}{1-\beta^2}\right)\right\}-1 \\ 1+\delta_x^2=\left(1+\delta^2\right)\exp\left(\dfrac{\sigma^2}{1-\beta^2}\right) \\ \rho_k=\left\{\exp\left(\dfrac{\sigma^2\beta^k}{1-\beta^2}\right)-1\right\}\Big/\left(1+\delta^2\right)\left\{\exp\left(\dfrac{\sigma^2}{1-\beta^2}\right)-1\right\} \\ \approx\dfrac{\sigma^2\beta^k/\left(1-\beta^2\right)}{\left(1+\delta^2\right)\left\{\exp\left(\dfrac{\sigma^2}{1-\beta^2}\right)-1\right\}}\approx\beta\rho_{k-1} \end{cases} \tag{2}$$

其中：$\mu=E(\varepsilon_i)$，$k=E\left(\varepsilon_i^2\right)/\mu$，$\delta_x$ 表示 x_i 离散化程度(标准差与均值的比率)，δ 表示 ε_i 离散化程度。

1.2 状态空间模型[4][5]

传统的线性高斯状态空间模型一般有多种表示，在这里我们使用以下形式：

$$\alpha_{t+1}=d_t+T_t\alpha_t+R_t\eta_t \quad (t=0,\cdots T) \tag{3}$$

$$y_t=c_t+Z_t\alpha_t+e_t \quad (t=1,\cdots,T) \tag{4}$$

其中，y_t 为一个可观测的 $k\times1$ 维向量，称为观测向量；α_t 为 $m\times1$ 维状态向量，不可观测；e_t 表示 $k\times1$ 向量，且 $e_t\sim N(0,H_t)$；R_t 表示 $m\times g$ 矩阵，η_t 表示 $g\times1$ 向量，且 $\eta_t\sim N(0,Q_t)$；$\alpha_1\sim N\left(\alpha_{1|0},\Sigma_{1|0}\right)$。状态方程(3)具有马尔可夫结构，可以有效的描绘时间序列 y_t 的序列相关结构；状态方程中的矩阵 d_t,T_t,R_t,Q_t 与观测方程中的矩阵 c_t,Z_t,H_t 一般假定为非随机的。因此，尽管它们能随时间改变，但是都是可以预先确定的。

在给定状态初值 $\alpha_{1|0}$ 及 $\Sigma_{1|0}$ 的情况下有Kalman滤波公式为：

$$\begin{cases} v_t=y_t-c_t-Z_t\alpha_{t|t-1}, V_t=Z_t\Sigma_{t|t-1}Z'_t+H_t \\ K_t=T_t\Sigma_{t|t-1}Z'_tV_t^{-1},\ L_t=T_t-K_tZ_t \\ \alpha_{t+1|t}=d_t+T_t\alpha_{t|t-1}+K_tv_t \\ \sum_{t+1|t}=T_t\sum_{t|t-1}L'_t+R_tQ_tR'_t \end{cases} \tag{5}$$

当线性高斯观测方程(3)被一般观测条件密度 $P\left(y|\alpha,\psi\right)$ 所代替，而状态转移方程(4)维持它最初的线性高斯形式时，将非高斯状态空间模型设定为：

$$P\{y_t|\alpha_1,\cdots,\alpha_t,y_1,\cdots,y_{t-1}\}=P\{y_t|Z_t\alpha_t\} \tag{6}$$

$$\alpha_{t+1}=T_t\alpha_t+R_t\eta_t,\ \eta_t\sim p(\eta_t) \tag{7}$$

其中，η_t 无序列相关。所谓非高斯，主要就是指无论 $P\{y_t|Z_t\alpha_t\}$ 或 $p(\eta_t)$ 都是非高斯分布。为了方便，我们记 $Z_t\alpha_t=\theta_t$。对于 $P\{y_t|\theta_t\}$ 的非高斯密度我们考虑以下两种情况。

①观测值来自指数族分布，即：

$$P\{y_t|\ \theta_t\}=\exp\{y'_t\theta_t-b_t(\theta_t)+c_t(y_t)\} \tag{8}$$

其中，$b_t(\theta_t)$ 二阶可微且 $c_t(y_t)$ 仅只是 y_t 的函数。

②观测值由下式产生：

$$y_t=\theta_t+e_t,\ e_t\sim p(e_t) \tag{9}$$

其中，$p(e_t)$ 为非高斯分布且 e_t 无序列相关。

2 模型估计方法[2][6]~[9]

在非高斯状态空间模型中令 $y=\left(y'_1,\cdots,y'_t\right)'$，$\alpha=\left(\alpha'_1,\cdots,\alpha'_t\right)'$，设 ψ 为未知的参数向量。通过蒙特卡罗似然函数而得到的极大似然估计记为 $\hat{\psi}$。对于给定的 ψ，令 $g\left(\alpha|\ y\right)$ 为重要密度(如高斯密度)，它与模型中的 $p\left(\alpha|\ y\right)$ 应尽可能接近。设 $\bar{x}=E\left[x(\alpha)|\ y\right]$ 表在给定观测向量 y 的条件下，α 的任意函数 $x(\alpha)$ 的条件均值，则

$$\bar{x}=\int x(\alpha)p\left(\alpha|\ y\right)d\alpha=\int x(\alpha)\frac{p\left(\alpha|\ y\right)}{g\left(\alpha|\ y\right)}g\left(\alpha|\ y\right)d\alpha$$

$$=E_g\left[x(\alpha)\frac{p\left(\alpha|\ y\right)}{g\left(\alpha|\ y\right)}\right] \tag{10}$$

其中，E_g 表重要密度 $g\left(\alpha|\ y\right)$ 的期望。

虽然 $p\left(\alpha|\ y\right)$ 与 $g\left(\alpha|\ y\right)$ 较难求，但联合密度 $p(\alpha,y)$ 与 $g(\alpha,y)$ 较容易求得。将 $p\left(\alpha|\ y\right)=p(\alpha,y)/p(y)$ 及 $g\left(\alpha|\ y\right)=g(\alpha,y)/g(y)$ 代入(10)式，得：

$$\bar{x}=\frac{g(y)}{p(y)}E_g\left[x(\alpha)\frac{p(\alpha,y)}{g(\alpha,y)}\right] \tag{11}$$

在(10)式中令 $x(\alpha)=1$，解得：

$$1=\frac{g(y)}{p(y)}E_g\left[\frac{p(\alpha,y)}{g(\alpha,y)}\right] \tag{12}$$

代入(11)得：

$$\bar{x}=\frac{E_g\left[x(\alpha)w(\alpha,y)\right]}{E_g\left[w(\alpha,y)\right]},\quad w(\alpha,y)=\frac{p(\alpha,y)}{g(\alpha,y)} \tag{13}$$

该公式可以用于估计条件协方差，条件密度，分布函数等。理论上我们可以获得 $\bar{x}$ 的蒙特卡罗估计 $\hat{x}$，从密度为 $g\left(\alpha|\ y\right)$ 的分布中抽取一列独立样本 $\alpha^{(1)},\cdots,\alpha^{(N)}$，则

$$\hat{x}=\sum_{i=1}^{N}x_iw_i\Big/\sum_{i=1}^{N}w_i \tag{14}$$

其中，$x_i=x\left(\alpha^{(i)}\right)$，$w_i=w\left(\alpha^{(i)},y\right)$

因为样本的独立性,所以在原假设下有 $\hat{x}\xrightarrow{p}\bar{x}(N\to\infty)$. 然而,样本估计在数值上是无效的,我们将在后面重新定义。当观测值是非高斯的,但状态向量由线性高斯模型(7)生成. 于是 $p(\alpha)=g(\alpha)$,所以:

$$\frac{p(\alpha,y)}{g(\alpha,y)}=\frac{p(\alpha)p(y|\alpha)}{g(\alpha)g(y|\alpha)}=\frac{p(y|\alpha)}{g(y|\alpha)}=\frac{p(y|\theta)}{g(y|\theta)}$$

式(13)成为样本公式:

$$\bar{x}=\frac{E_g\left[x(\alpha)w^*(\theta,y)\right]}{E_g\left[w^*(\theta,y)\right]},\quad w^*(\theta,y)=\frac{p(y|\theta)}{g(y|\theta)}$$

其估计 $\hat{x}$ 显然与(14)式类似。

对于 $p(\alpha|y)$,我们通过使用其条件众数 $\hat{\alpha}$ 来构造近似线性高斯模型,用以生成重要高斯密度以便模拟。实证研究的经验表明众数与均值没有太明显的差别,使用众数不足的地方是不能一起估计误差的协方差矩阵。

令 $g(\alpha|y)$, $g(\alpha,y)$ 分别表示由线性高斯模型:

$$y_t=Z_t\alpha_t+e_t\ ,\ e_t\sim N(0,H_t) \tag{15}$$

$$\alpha_{t+1}=T_t\alpha_t+R_t\eta_t,\eta_t\sim N(0,Q_t) \tag{16}$$

生成的条件及联合密度,而令 $p(\alpha|y)$, $p(\alpha,y)$ 分别表示由非高斯模型(6)、(7)生成的条件及联合密度。我们将选择适当的 H_t 和 Q_t ,使得 $g(\alpha|y)$ 和 $p(\alpha|y)$ 具有相同的众数 $\hat{\alpha}$ (因为近似模型仅依赖于均值和方差,且是唯一的,故 $p(\alpha,y)$ 的多峰性可暂不考虑)。

首先分析高斯模型, $\hat{\alpha}$ 应为向量方程 $\partial\left[\log\{g(\alpha|y)\}\right/\partial\alpha\right]=0$ 的解。

$$\text{因为}\frac{\partial\left[\log\{g(\alpha|y)\}\right]}{\partial\alpha}=\frac{\partial\left[\log\{g(\alpha,y)\}-\log\{g(y)\}\right]}{\partial\alpha}$$

$$=\frac{\partial\left[\log\{g(\alpha,y)\}\right]}{\partial\alpha}=0,\text{所以 }\hat{\alpha}\text{ 也是向量方程}$$

$\partial\left[\log\{g(\alpha,y)\}\right]/\partial\alpha=0$ 的解。

又因为 R_t 包含正定矩阵 I_m 的 r 列,所以 $\eta_t=R'_t(\alpha_t-T\alpha_{t-1})$ 。由于

$$g(\alpha,y)=g(\alpha_1)\prod_{t=1}^{n}g(\eta_t)g(y|\alpha_t)$$

$$\log\{g(\alpha,y)\}=\text{const}\,ant-\frac{1}{2}\sum_{t=1}^{n}(\alpha_t-T_t\alpha_{t-1})'$$

$$R_tQ_t^{-1}R_t'(\alpha_t-T_t\alpha_{t-1})-\frac{1}{2}\sum_{t=1}^{n}(y_t-Z_t\alpha_t)'H_t^{-1}(y_t-Z_t\alpha_t)$$

对上式关于 α_t 求微分并令结果等于0,得:

$$-R_tQ_t^{-1}R_t'(\alpha_t-T_t\alpha_{t-1})+s_tT'_{t+1}R_{t+1}Q_{t+1}^{-1}R'_{t+1}$$
$$(\alpha_{t+1}-T_{t+1}\alpha_t)+Z'_tH_t^{-1}(y_t-Z_t\alpha_t)=0 \tag{17}$$

其中, $s_t=1(t=1,2,\cdots,n-1)$, $s_n=0$ 。

求解(17)式便得到条件众数 $\hat{\alpha}$.实际由于 $g(\alpha,y)$ 为高斯密度,众数等于均值,所以 $\hat{\alpha}$ 可以通过Kalman 滤波和平滑求得。假设模型方程也是规范的,则 $p(\alpha|y)$ 的众数 $\hat{\alpha}$ 也是向量方程 $\partial\left[\log\{p(\alpha|y)\}\right]/\partial\alpha=0$ 的解,同理 $\hat{\alpha}$ 也是向量方程 $\partial\left[\log\{p(\alpha,y)\}\right]/\partial\alpha=0$ 的解。

设 $q_t(\eta_t)=-\log\{p(\eta_t)\}$, $h_t(y_t|\theta_t)=-\log\{p(y_t|\theta_t)\}$,则

$$\log\{p(\alpha,y)\}=\text{const}\,ant-\sum_{t=1}^{n}\{q_t(\eta_t)+h_t(y_t|\theta_t)\}$$

仍然有 $\eta_t=R'_t(\alpha_t-T\alpha_{t-1})$,于是 $\hat{\alpha}$ 由下式解出

$$\frac{\partial\left[\log\{p(\alpha,y)\}\right]}{\partial\alpha_t}=-R_t\frac{\partial q(\eta_t)}{\partial\eta_t}+s_tT'_{t+1}R_{t+1}\frac{\partial q_{t+1}}{\partial\eta_{t+1}}$$
$$-Z'_t\frac{\partial h_t(y_t|\theta_t)}{\partial\theta_t}=0 \tag{18}$$

其中, $s_t=1(t=1,2,\cdots,n-1)$, $s_n=0$ 。我们将利用迭代算法解出该等式,只要在每一步将其线性化便可得到如(17)式的形式,于是便可利用Kalman 滤波和平滑算法进行求解。收敛速度通常很快的,一般最多10次迭代即可。

对于观测密度为非高斯的情况,我们采取线性化这一常见方法。设 α 的一组实验值为 $\tilde{\alpha}=(\tilde{\alpha}'_1,\cdots,\tilde{\alpha}'_n)'$, $\tilde{\theta}_t=Z_t\tilde{\alpha}_t$,并定义

$$\dot{h}_t=\frac{\partial h_t(y_t|\theta_t)}{\partial\theta_t}\Big|_{\theta_t=\tilde{\theta}_t},\quad \ddot{h}_t=\frac{\partial^2h_t(y_t|\theta_t)}{\partial\theta_t\partial\theta'_t}\Big|_{\theta_t=\tilde{\theta}_t} \tag{19}$$

将 $\frac{\partial h_t(y_t|\theta_t)}{\partial\theta_t}$ 在 $\tilde{\theta}_t$ 处用一阶泰勒展式近似,即:

$$\frac{\partial h_t(y_t|\theta_t)}{\partial\theta_t}=\dot{h}_t+\ddot{h}_t(\theta_t-\tilde{\theta}_t) \tag{20}$$

将(20)式代入(18)式的最后一项,即:

$$-Z'(\dot{h}_t+\ddot{h}_t\theta_t-\ddot{h}_t\tilde{\theta}_t)=Z'_t\ddot{h}_t(\tilde{\theta}_t-\ddot{h}_t^{-1}\dot{h}_t-\theta_t)$$

令 $\tilde{H}_t=\ddot{h}_t^{-1}$, $\tilde{y}_t=\tilde{\theta}_t-\ddot{h}_t^{-1}\dot{h}_t$,则(18)式的最后一项变为 $Z'_t\tilde{H}_t^{-1}(\tilde{y}_t-\theta_t)$,这样(18)式就转化成了形如(17)式的线性形式。在给定一组新实验值的情况下通过Kalman滤波和平滑可解出 α ,且让该过程反复进行直至收敛。对收敛后的 α 和 θ 将其分别记为 $\hat{\alpha}$ 和 $\hat{\theta}$ 即可。

对SCD模型(1)式中的第一式取对数变换后变换成状态空间形式为:

$$\begin{cases}Z_i=\ln(x_i)=c+\psi_i+\eta_i\\ \psi_i=\omega+\beta\psi_{i-1}+u_i,(|\beta|<1)\end{cases} \tag{21}$$

其中, $\eta_i=\ln\varepsilon_i-c$, $c=E[\ln\varepsilon_i]$,在模型假定下 $\psi_i\sim N\left(\frac{\omega}{1-\beta},\frac{\sigma^2}{1-\beta^2}\right)$ 。由于SCD模型中的扰动项 ε_i 不服从对数高斯分布,故此状态空间形式属于非高斯状态空间模型。若要对此模型进行QML估计,便假设扰动项服从高斯分布,即 $\eta_i\sim N(0,\text{Var}[\ln\varepsilon_i])$,QML估计虽然是一致的但与ML估计相比并非有效。

利用本文非高斯状态空间模型框架下的滤波公式对

SCD模型进行估计，首先对非高斯分布进行线性化，设定出 $\tilde{H}_i$ 及 $\tilde{x}_i$ 。根据常见的 ε_i 的分布可列出相应的条件密度见表1，根据（19）、（21）两式，及 $\tilde{H}_i=\ddot{h}_i^{-1}$ ，$\tilde{x}_i=\psi_i-\tilde{H}_i\dot{h}_i^{-1}$，相应的 $\tilde{H}_i$ 及 $\tilde{x}_i$ 见表2。

表1　　持续期的条件分布与相应的条件密度

条件分布	条件密度 $p(x_i\|\psi_i)$
指数分布	$\exp(-\psi_i)\exp\{-x_i\exp(-\psi_i)\}$
Weibull分布	$\gamma\left(\frac{\Gamma(1+\gamma^{-1})}{\exp(\psi_i)}\right)^{\gamma}x_i^{\gamma-1}\exp\left\{-\left[-\frac{x_i}{\exp(\psi_i)}\Gamma(1+\gamma^{-1})\right]^{\gamma}\right\}$
Gamma分布	$\left(\frac{\nu}{\exp(\psi_i)}\right)^{\nu}x_i^{\nu-1}\exp\left\{-\frac{x_i\nu}{\exp(\psi_i)}\right\}\Big/\Gamma(\nu)$

表2　　相应分布所对应的 $\tilde{H}_i$ 及 $\tilde{x}_i$

条件分布	$\tilde{H}_i$	$\tilde{x}_i$
指数分布	$\tilde{H}_i=\tilde{x}_i\exp(\psi_i)$	$\tilde{x}_i=\psi_i-x_i^{-1}\exp(\psi_i)+1$
Weibull分布	$\tilde{H}_i=\left[\frac{x_i}{\exp(\psi_i)}\Gamma(1+\gamma^{-1})\right]^{-\lambda}$	$\tilde{x}_i=\psi_i-\gamma\left[\frac{x_i}{\exp(\psi_i)}\Gamma(1+\gamma^{-1})\right]^{-\gamma}$
Gamma分布	$\tilde{H}_i=\frac{\exp(\psi_i)}{\nu x_i}$	$\tilde{x}_i=\psi_i-\frac{\exp(\psi_i)}{x_i}+1$

3　模拟结果

为了更好的比较非高斯状态空间模型下对SCD模型所做估计（记为KML）与QML估计。本文设计了几组模拟数据实验，考虑了几种不同情况的参数值。由SCD数据生成过程DGP产生三组模拟样本，样本数分别为500,2000,10000，然后将两种估计的估计值及相应的标准差，相对标准差作比较。由于关于SCD模型估计方面的文献大多考虑扰动项服从Weibull分布，我们主要考虑扰动项服从Gamma分布这一情况。DGP由（1）式及 $p(x_i|\psi_i,\nu)=\left(\frac{\nu}{\exp(\psi_i)}\right)^{\nu}x_i^{\nu-1}\exp\left\{-\frac{x_i\nu}{\exp(\psi_i)}\right\}\Big/\Gamma(\nu)$ 组成，用于生成模拟数据的参数定义为：$\omega=0.02$ ，$\beta=0.9$和0.95 ，$\sigma=0.2$和0.1 ，$\nu=0.9$ 。从表3～表5可以看出，估计值与真实值相差很小，且随着样本数的增加，估计值与真实值之间的差距变小，标准差也随着样本容量的增加而减小。

表3　　KML与QML的模拟估计结果（N=500）

500	真实值	估计值		标准差		相对标准差
		KML	QML	KML	QML	QML/KML
ω	0.020	0.0247	−0.0023	0.0027	0.0066	2.4444
β	0.900	0.8325	0.8152	0.0157	0.0219	1.3949
σ	0.100	0.1334	0.1537	0.012	0.0135	1.1250
ν	0.900	0.8706	0.8052	0.061	0.1637	2.6836
ω	0.020	0.0235	0.0384	0.0025	0.0140	5.6000
β	0.900	0.7442	0.6791	0.0335	0.4014	11.982
σ	0.200	0.2114	0.5753	0.0326	0.2017	6.1871
ν	0.900	0.8734	0.8146	0.0039	0.3657	39.3226
ω	0.020	0.0283	−0.0468	0.0062	0.1928	31.097
β	0.950	0.6586	−0.0961	0.1649	1.3527	8.2032
σ	0.100	0.2966	0.5378	0.3224	0.9812	3.0434
ν	0.900	0.7452	0.7152	1.2137	1.653	1.3620
ω	0.020	0.0269	0.1473	0.0054	0.0835	15.463
β	0.950	0.8721	0.8023	0.0289	0.1131	3.9135
σ	0.200	0.1547	0.1683	0.0053	0.0156	2.9434
ν	0.900	0.8563	0.8332	0.0058	0.2451	42.2586

表4　　KML与QML的模拟估计结果（N=2000）

2000	真实值	估计值		标准差		相对标准差
		KML	QML	KML	QML	QML/KML
ω	0.020	0.0215	0.0203	0.0004	0.0009	2.2500
β	0.900	0.8927	0.8841	0.0013	0.0023	1.7692
σ	0.100	0.1152	0.1191	0.0365	0.0670	1.8356
ν	0.900	0.8533	0.8043	0.0230	0.0540	2.3478
ω	0.020	0.0234	0.0597	0.0017	0.0480	28.2353
β	0.900	0.8590	1.2427	0.4163	1.3290	3.1924
σ	0.200	0.2184	0.5416	0.0626	0.8305	13.2668
ν	0.900	0.8518	0.7906	0.0021	0.0535	25.4762
ω	0.020	0.0035	−0.0587	0.0511	0.8350	16.3405
β	0.950	0.7786	0.3463	0.0834	1.3524	16.2158
σ	0.100	0.1074	0.5389	0.0234	0.4288	18.3248
ν	0.900	0.7586	0.7369	0.8933	0.9899	1.1081
ω	0.020	0.0225	0.0156	0.0015	0.0052	3.4667
β	0.950	0.8896	0.8741	0.0032	0.0044	1.3750
σ	0.200	0.2158	0.1139	0.0026	0.0073	2.8077
ν	0.900	0.8337	0.8239	0.0148	0.0212	1.4324

表5　　KML与QML的模拟估计结果（N=10000）

10000	真实值	估计值		标准差		相对标准差
		KML	QML	KML	QML	QML/KML
ω	0.020	0.0201	0.0202	0.0000	0.0000	1.0000
β	0.900	0.9005	0.8910	0.0006	0.0010	1.6667
σ	0.100	0.1030	0.1120	0.0015	0.0027	1.8000
ν	0.900	0.8947	0.8599	0.0023	0.0071	3.0870
ω	0.020	0.0201	0.0201	0.0003	0.0004	1.3333
β	0.900	0.9008	0.8892	0.0001	0.0004	4.0000
σ	0.200	0.1988	0.1974	0.0002	0.0007	3.5000
ν	0.900	0.9001	0.9003	0.0001	0.0002	2.0000
ω	0.020	0.0202	0.0189	0.0001	0.0002	2.0000
β	0.950	0.9480	0.9368	0.0004	0.0005	1.2500
σ	0.100	0.1010	0.1354	0.0026	0.0672	25.8462
ν	0.900	0.8897	0.8679	0.0002	0.0008	4.0000
ω	0.020	0.0200	0.0201	0.0000	0.0001	1.0000
β	0.950	0.9498	0.9613	0.0061	0.0061	1.0000
σ	0.200	0.2000	0.2003	0.0000	0.0001	1.0000
ν	0.900	0.9001	0.8997	0.0001	0.0002	2.0000

从表中可以看出，本文所采用的方法得到的估计值都比QML估计更加接近于真实值，且标准差都比QML的来得小。可以说在给定抽样密度下，KML估计比QML更精确。

4　结论

本文研究了非线性非高斯情况下的状态空间分析框架，对于观测值的条件分布及扰动项的非高斯分布情况均作了考虑。针对金融时间序列的常见分布，考虑了观测变量来自指数族分布以及两个方程中扰动项的重尾分布及二阶可导的情况，研究方法主要是基于一阶泰勒展式来进行近似线性化，转化到线性状态空间框架下来研究可用的Kalman滤波公式。并针对当前比较热门的高频超高频金融计量研究，利用非高斯情况下的状态空间模型下的滤波研究了随机条件持续

超出损失保险的纯保费估计和风险度量

欧阳资生

（湖南商学院 金融学院，长沙 410205）

摘　要：文章引入负二项分布作为保单组合的索赔次数的分布，全Paretian分布作为索赔额的分布，研究了超出损失保险保单组合的纯保费的估计方法，导出了最大可能损失的估计式，并利用火灾保险索赔数据进行了实证分析。

关键词：火灾保险；纯保费；全Paretian模型；最大可能损失

中图分类号：F222.3　　**文献标识码：**A　　**文章编号：**1002-6487（2011）15-0021-03

0　引言

根据瑞士Sigma杂志统计，自1990年以来全球巨灾损失(包括自然灾害和人为灾祸)发生变得更加频繁和严重。巨大的保险损失给国际保险业带来了新的挑战，严重威胁着保单持有者的利益。单纯的依靠保险业自身的实力已经无法应对这些巨灾风险，保险公司的承保能力急剧下降，保险公司和保险监管部门亟待研究制定新的巨灾保险费率的方法以解决这些保险公司的赔付能力。

当然，保险业发展至今日，其风险转移的方式已多样化，如再保险、保险衍生证券等都已经成为很好的风险转移的方式。在再保险中，精算师研究的实际上就是超出损失保险(Excess-of-Loss)(简称 XL)的纯保费问题。在 XL 中，被保险人只能对超出某一固定值(门限值) 的损失部分提出索赔要求。这时，纯保费就是下一个时期内总索赔数目的期望值。如果设 S_N 为下一个时期内总索赔数目，N 是下一个时期内索赔额超出某一固定值 u 的索赔次数，Z_i 为超出 u 的索赔额，X_i 表示索赔额超出固定值 u 的该张保单的索赔额。那么，$Z_i = X_i - u$ 。而下一期总索赔数目就是

$$S_N = \sum_{i \le N} Z_i \tag{1}$$

因此纯保费就是 $E[S_N]$ 。

对于式(1)的研究，至少涉及到两个方面，一是索赔次数 N 的估计问题，目前文献中索赔次数大多采用Poisson过程来描述。事实上，如果 N 为服从一参数为 λ 的Poisson过

基金项目：教育部人文社会科学基金资助项目（09YJA910003）和湖南省软科学资助项目(2009ZK4022)

作者简介：欧阳资生（ 1967-），男，湖南邵阳人，博士后，教授，研究方向：风险管理与精算学。

期SCD模型的估计，从经验分析中考虑了SCD模型中的扰动项服从伽马分布这一情况。本文研究的非高斯状态空间模型下的Kalman滤波对SCD模型进行的估计及QML估计，均表明KML估计比QML估计更精确，标准更小。相信随着人们对状态空间理论的研究，状态空间理论一定能得到更大的发展，从而解决金融领域中更多非线性、非高斯，甚至是产生结构性变化的问题。

参考文献：

[1]R.F.Engle， J.R， Russell. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data[J].Econometrica, 1998,（66）.

[2]Luc Bauwens, David Veredas. The Stochastic Conditional Duration Model: a Latent Variable Model for the Analysis of Financial Durations [J]. Journal of Econometrics，2004，（119）.

[3]Chris M. Strickland, Catherine S. Forbes，Gael M. Martin. Bayesian Analysis of the Stochastic Conditional Duration Model[J]. Computational Statistics & Data Analy Sis, 2006,（50）.

[4]S.J.Koopman, J.Durbin. Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Model[J]. Journal of the American Statistical Association, 1997,（92）.

[5]Ruey S，Tsay. Analysis of Financial Time Series（2nd Edition）[M]. Hoboken: John Wiley@Sons,Inc, 2005 .

[6]J.Durbin, S.J.Koopman. Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models[J]. Biometrika, 1997, 84(3).

[7]N.Shephard, M.K.Pitt. Likelihood Analysis of Non-Gaussian Measurement Time Series[J]. Biometrika,1997,（84）.

[8]J.Durbin, S.J.Koopman. Time Series Analysis of Non-Gaussian Observations Based on State Space Models From Both Classical and Bayesian Perspectives[J]. Journal of the Royal Statistical Society,2000, Series B, 62(1) 6.

[9]J.Durbin, S.J.Koopman. A Simple and Efficient Simulation Smoother for State Space Time Series Analysis[J]. Biometrika,2002,89(3).

[10]耿克红，张世英. SCD模型与ACD模型比较研究[J].管理学报，2008，5(1).

（责任编辑/亦　民）